DOCUMENT RESUME

ED 078 056 TM 002 872

AUTHOR Archer, Claud O.; Jennrich, Robert I.
TITLE Standard Errors for Rotated Factor Loadings.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RB-73-17
PUB DATE Mar 73
NOTE 23p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Algorithms; Data Analysis; *Factor Analysis; Matrices; Measurement Techniques; Orthogonal Rotation; *Statistical Analysis; Technical Reports

ABSTRACT

Beginning with the results of Girschick on the asymptotic distribution of principal component loadings and those of Lawley on the distribution of unrotated maximum likelihood factor loadings, the asymptotic distributions of the corresponding analytically rotated loadings is obtained. The principal difficulty is the fact that the transformation matrix which produces the rotation is usually itself a function of the data. The approach is to use implicit differentiation to find the partial derivatives of an arbitrary orthogonal rotation algorithm. Specific details are given for the orthomax algorithms and an example involving maximum likelihood estimation and varimax rotation is presented. (Author)

RB-73-17

RESEARCH BULLETIN

STANDARD ERRORS FOR ROTATED FACTOR LOADINGS

Claud O. Archer
Naval Test Center

and

Robert I. Jennrich
University of California at Los Angeles
and Educational Testing Service

Educational Testing Service
Princeton, New Jersey
March 1973

STANDARD ERRORS FOR ROTATED FACTOR LOADINGS

Abstract

Beginning with the results of Girshick on the asymptotic distribution of principal component loadings and those of Lawley on the distribution of unrotated maximum likelihood factor loadings, the asymptotic distributions of the corresponding analytically rotated loadings is obtained. The principal difficulty is the fact that the transformation matrix which produces the rotation is usually itself a function of the data. The approach is to use implicit differentiation to find the partial derivatives of an arbitrary orthogonal rotation algorithm. Specific details are given for the orthomax algorithms and an example involving maximum likelihood estimation and varimax rotation is presented.

STANDARD ERRORS FOR ROTATED FACTOR LOADINGS*

1. Introduction

While its proponents have never questioned its importance, many statisticians are surprised to learn that factor analysis is one of the most popular methods of statistical investigation. An extensive computer usage survey at UCLA found regression analysis, discriminant analysis, and factor analysis the three most popular statistical methodologies. Informal inquiries at other institutions indicate that more often than not, factor analysis ranks in the top three. Computer programs for factor analysis are unusual, however, in that they give no standard errors for the estimates they produce. This is due in large measure to the fact that until now formulas for the standard errors of estimates of rotated factor loadings, the estimates which constitute the primary output of standard factor analysis programs, have not been produced. In two important papers Lawley [1953, 1967] identified the asymptotic standard errors of the unrotated loadings produced in maximum likelihood factor analysis. Similar results for principal components analysis were given some time ago by Girshick [1939]. The difficulty in extending these results to the case of rotated loadings is that the transformation matrix T which produces the rotation is usually derived from the data. It is perhaps not surprising that Lawley and Maxwell [1971] state, "It would be almost impossible to take sampling errors in the elements of T into account. The only course is, therefore, to ignore them in the hope that they are relatively small."

---

*This research was supported in part by NIH Grant FR-3. The authors are grateful to Mrs. Dorothy T. Thayer who implemented the algorithms discussed here as well as those of Lawley and Maxwell. We are particularly indebted to Dr. Michael Browne for convincing us of the significance of this work and for helping to guide its development.

We shall show that these sampling errors can in fact be taken into account. This is important because Wexler [1968] has produced some evidence to indicate they cannot be safely ignored.

We begin with a p by k factor loading matrix $A = (\alpha_{ir})$ and an asymptotically normal estimate $\hat{A} = (\hat{\alpha}_{ir})$. By this we mean that as the size n of the sample on which $\hat{A}$ is based approaches infinity, the distribution of $\sqrt{n}\,(\hat{A} - A)$ approaches multivariate normal with mean zero. The maximum likelihood estimates in the classical factor analysis model and the principal components estimates in principal components analysis both have this property. We are not concerned at this point with specifically which estimates are being considered, but rather with the effect of an orthogonal rotation algorithm on the asymptotic distribution of $\hat{A}$. We have in mind algorithms such as quartimax, varimax, and equimax, but for the present let h denote an arbitrary orthogonal rotation algorithm. Specifically h is a function which maps an arbitrary p by k matrix X into a p by k matrix $Y = XT$ where T is an orthogonal matrix whose value may, and generally will, depend on X. We are interested in the asymptotic distribution of $\hat{\Lambda} = h(\hat{A}) = (\hat{\lambda}_{ir})$. In particular if $\Lambda = h(A) = (\lambda_{ir})$ we would like to conclude that $\sqrt{n}\,(\hat{\Lambda} - \Lambda)$ is asmptotically normally distributed and to find its asymptotic covariance matrix. This we shall do.

2. The Asymptotic Distribution of Orthogonally Roteted Loadings

In principle at least our task is quite simple. Let dh be the differential of h at A. Then

$$(1) \qquad \sqrt{n}\,(\hat{\Lambda} - \Lambda) \overset{a}{=} dh(\sqrt{n}\,(\hat{A} - A))$$

where " $\overset{a}{=}$ " is read "is asymptotically equal to" and means that the difference between the left and right sides of (1) approaches zero in probability as $n \to \infty$. This is the basis of the δ-method as discussed by Rao [1965, p. 321]. Since dh is a linear transformation, and $\hat{A}$ is an asymptotically normal estimator of A, $\hat{\Lambda}$ is an asymptotically normal estimator of Λ whose asymptotic covariance matrix may be obtained from dh and the asymptotic covariance matrix of $\hat{A}$. To be more explicit, let the differential of the relation $\Lambda = h(A) = (h_{ir})$ be expressed by a formula of the form

$$(2) \qquad d\lambda_{ir} = \Sigma_{js} \frac{\partial h_{ir}}{\partial \alpha_{js}} d\alpha_{js} \quad .$$

Then the asymptotic covariances of the $\hat{\lambda}_{ir}$ may be expressed in terms of those of the $\hat{\alpha}_{ir}$ by means of the formula

$$(3) \qquad \operatorname{acov}(\hat{\lambda}_{ir}, \hat{\lambda}_{js}) = \Sigma_{mnuv} \frac{\partial h_{ir}}{\partial \alpha_{mu}} \operatorname{acov}(\hat{\alpha}_{mu}, \hat{\alpha}_{nv}) \frac{\partial h_{js}}{\partial \alpha_{nv}} \quad .$$

For rotation algorithms of interest, quartimax, varimax, and equimax, it is difficult to find dh or, equivalently, the partial derivatives $\partial h_{ir}/\partial \alpha_{js}$ directly. Our approach is to use implicit differentiation. Suppose that

$$(4) \qquad \psi(\Lambda) = 0$$

is a $k(k-1)/2$ dimensional constraint which is satisfied whenever Λ is an h-rotation of A no matter what the value of A. Constraints of this form for rotation algorithms of interest will be found in Section 5. By differentiating the relations

$$(5) \qquad \Lambda = AT \quad , \quad T'T = I \quad , \quad \psi(\Lambda) = 0$$

one obtains

$$(6) \qquad d\Lambda = dAT + AdT$$

$$(7) \qquad dT'T + T'dT = 0$$

$$(8) \qquad d\psi\,(d\Lambda) = 0$$

where $d\psi$ denotes the differential of ψ at Λ. It follows from (7) that $T'dT$ is a skew-symmetric k by k matrix. Let $\mathcal{K}$ denote the space of all such matrices. It has dimension $k(k-1)/2$. Moreover, for each $K \in \mathcal{K}$ let

$$(9) \qquad L(K) = d\psi(\Lambda K) .$$

Then L is a linear transformation from a $k(k-1)/2$ dimensional space into a $k(k-1)/2$ dimensional space which we assume is invertible. This is usually the case for constraint functions ψ of interest. Since $T'dT$ is skew-symmetric it is in the domain of L and using (9) and (5) in order gives

$$(10) \qquad L(T'dT) = d\psi(\Lambda T'dT) = d\psi(AdT) \quad .$$

Substituting (6) into (8) and using the linearity of $d\psi$ shows that $d\psi(AdT) = -d\psi(dAT)$ so that from (10),

(11) $$L(T'dT) = d\psi(AdT) = -d\psi(dAT) \quad .$$

Thus

(12) $$T'dT = L^{-1}[d\psi(AdT)] = -L^{-1}[d\psi(dAT)] \quad .$$

Multiplying on the left by AT and using (5) and (6) gives the basic relation

(13) $$d\Lambda = dAT - \Lambda L^{-1}[d\psi(dAT)]$$

which expresses $d\Lambda$ in terms of dA and defines the differential of h at A. It also defines the required partial derivatives of h. All that is needed for a particular rotation algorithm is to find an appropriate constraint function ψ and to recover the partial derivatives $\partial h_{ir}/\partial\alpha_{js}$ from (13). The first task will be relatively easy. The second is a little harder.

3. Constraints for an Orthogonal Algorithm

Orthogonal rotation algorithms are designed to optimize a criterion:

(14) $$Q = Q(\Lambda) = Q(AT)$$

over all orthogonal k by k matrices T. The resulting $\Lambda = AT$ is called the Q-rotation of A. In the case of quartimax, varimax, and equimax rotation, Q is a quartic function of Λ. In target rotation, on the other hand, Q is quadratic. There is, however, no need to specialize at this point. An arbitrary Q is considered here and in the next section.

Assume that $\Lambda = AT$ optimizes Q and let dQ denote the differential of Q at Λ. It is necessary that

(15) $$dQ(AdT) = 0$$

for all dT which satisfy (7). Since $T'dT$ is skew-symmetric, it is necessary that

(16) $$dQ(\Lambda K) = 0$$

for all $K \in \mathcal{K}$. Let $K(r,s)$ be the elementary k by k skew-symmetric matrix which has the value 1 in row r and column s, the value -1 in row s and column r, and is zero elsewhere. Replacing K in (16) by $K(r,s)$ and writing the result in coordinate form gives:

(17) $$\psi_{rs} = \sum_{i=1}^{p} \left(\lambda_{ir} \frac{\partial Q}{\partial \lambda_{is}} - \lambda_{is} \frac{\partial Q}{\partial \lambda_{ir}}\right) = 0$$

for $1 \leq r,s \leq k$. These are the constraints which are needed. We observe that the matrix $\psi = (\psi_{rs})$ is skew-symmetric for arbitrary Λ.

One may view (17) slightly differently. Let $\frac{dQ}{d\Lambda}$ denote the p by k matrix $(\partial Q/\partial \lambda_{ir})$ of partial derivatives of Q. Then (17) says that $\Lambda' \frac{dQ}{d\Lambda}$ is symmetric. In the case of the quartimax criterion $\frac{dQ}{d\Lambda} = \Lambda^3$ where $\Lambda^3 = (\lambda_{ir}^3)$, and (17) demands that $\Lambda'\Lambda^3$ be symmetric. This furnishes a simple test for the convergence of a quartimax algorithm. Corresponding tests apply to other orthogonal algorithms.

4. The Differential of an Orthogonal Algorithm

We turn now to the problem of finding formulas for the partial derivatives of an orthogonal rotation algorithm h. Note first that the elementary skew-symmetric matrices

$$(18) \qquad K(u,v) \quad , \quad 1 \leq u < v \leq k$$

form a basis for $\mathcal{A}$. The $k(k-1)/2$ by $k(k-1)/2$ matrix (L_{ij}) of L as defined in (9) relative to this basis taken in lexicographic order is given by:

$$(19) \qquad L_{\ell(r,s),\ell(u,v)} = d\psi_{rs}(\Lambda K(u,v))$$

$$= \sum_{i=1}^{p} \left(\lambda_{iu} \frac{\partial\psi_{rs}}{\partial\lambda_{iv}} - \lambda_{iv} \frac{\partial\psi_{rs}}{\partial\lambda_{iu}}\right)$$

where

$$(20) \qquad \ell(r,s) = (r-1)(2k-r)/2 + s - r$$

for $1 \leq r < s \leq k$ and $1 \leq u < v \leq k$.

While it is not evident from our derivation and not essential to what follows, the matrix (L_{ij}) is in fact symmetric and nonnegative definite.* Under the assumption that L is nonsingular, (L_{ij}) is positive definite.

*It can be shown that (L_{ij}) is a matrix of second partial derivatives evaluated at the minimum of an appropriate function.

Let (L^{ij}) be the inverse of the matrix (L_{ij}) and let

$$(21) \qquad e_{iruv} = [\Lambda L^{-1}(K(u,v))]_{ir}$$

be the (i,r)-th component of the matrix $\Lambda L^{-1}(K(u,v))$. Using the fact that the (L^{ij}) is the matrix of L^{-1} with respect to the basis in (18),

$$(22) \qquad e_{iruv} = \sum_{t=1}^{r-1} \lambda_{it} L^{\ell(t,r),\ell(u,v)} - \sum_{t=r+1}^{k} \lambda_{it} L^{\ell(r,t),\ell(u,v)}$$

for $1 \leq i \leq p$, $1 \leq r \leq k$, and $1 \leq u < v \leq k$. The sums in (18) are zero when the lower limit exceeds the upper limit.

Finally, using (21), the basic relation (13) may be put in the coordinate form:

$$(23) \qquad d\lambda_{ir} = \sum_{s=1}^{k} d\alpha_{is} T_{sr} - \sum_{u<v} \sum_{jst} e_{iruv} \frac{\partial \psi_{uv}}{\partial \lambda_{jt}} d\alpha_{js} T_{st} \quad .$$

Reading the partial derivatives of h from this gives:

$$(24) \qquad \frac{\partial h_{ir}}{\partial \alpha_{js}} = \delta_{ij} T_{sr} - \sum_{u<v} \sum_{t} e_{iruv} \frac{\partial \psi_{uv}}{\partial \lambda_{jt}} T_{st}$$

for $1 \leq i,j \leq p$ and $1 \leq r,s \leq k$. Here δ_{ij} denotes the Kronecker delta.

In summary the partial derivatives of h may be computed from Λ, T, and a formula for Q as follows:

(i) Use Q and (17) to obtain formulas for the ψ_{rs}.

(ii) Compute the values of the partial derivatives $\partial\psi_{rs}/\partial\lambda_{it}$.

(iii) Using (19) form the matrix (L_{ij}) and invert it.

(iv) Compute the e_{iruv} using (21).

(v) Using (24) compute the partial derivatives of h.

5. Orthomax Algorithms

We turn now to the problem of finding specific formulas for the orthomax algorithms. These algorithms are designed to maximize the orthomax criterion:

$$Q = \frac{1}{4} \sum_{r=1}^{k} \left(\sum_{i=1}^{p} \lambda_{ir}^4 - \frac{\gamma}{p} \left(\sum_{i=1}^{k} \lambda_{ir}^2 \right)^2 \right) . \tag{25}$$

This becomes the quartimax, varimax, and equimax criterion when $\gamma = 0,1$, and $k/2$ respectively. Using (17) the corresponding constraint functions are:

$$\psi_{rs} = \sum_{i=1}^{p} \lambda_{ir}\lambda_{is}(\lambda_{ir}^2 - \lambda_{is}^2) - \frac{\gamma}{p} \sum_{i=1}^{p} \lambda_{ir}\lambda_{is} \sum_{i=1}^{p} (\lambda_{ir}^2 - \lambda_{is}^2) \tag{26}$$

for $1 \leq r,s \leq k$. Alternatively, these constraint functions may be found, in the $\gamma = 0,1$ cases at least, by setting the rotation angle φ equal to zero in the quartimax and varimax algorithms described by Harman [1967, p. 300 and p. 307].

The partial derivatives of the ψ_{rs} follow easily from (26). For $1 \leq i \leq p$ and $1 \leq r \neq s \leq k$ they are:

$$\begin{aligned} \frac{\partial\psi_{rs}}{\partial\lambda_{ir}} &= 3\lambda_{ir}^2\lambda_{is} - \lambda_{is}^3 - \frac{\gamma}{p} [\lambda_{is} \sum_{j=1}^{p} (\lambda_{jr}^2 - \lambda_{js}^2) + 2\lambda_{ir} \sum_{j=1}^{p} \lambda_{jr}\lambda_{js}] \\ \frac{\partial\psi_{rs}}{\partial\lambda_{is}} &= - \frac{\partial\psi_{sr}}{\partial\lambda_{is}} . \end{aligned} \tag{27}$$

For all other values of i , r , s , and t :

$$\frac{\partial \psi_{rs}}{\partial \lambda_{it}} = 0 \quad . \tag{28}$$

In summary the partial derivatives of an orthomax algorithm h may be computed from Λ and T as follows:

(i) Using (27) and (28) compute the partial derivatives of ψ .

(ii) Using (19) form (L_{ij}) and invert it.

(iii) Using (21) compute the c_{iruv} .

(iv) Use (24) to give the partial derivatives of h .

As observed earlier, we must assume that L as defined by (9) is nonsingular when Λ is an orthomax rotation of A . This needs to be an assumption for it is not always true. It is easy to show, for example, that if the orthomax criterion has the same value for every rotation of A , then $L \equiv 0$ and is clearly singular. In the two factor case this is the only way in which L can be singular. More generally, in the course of a simulation study the authors have looked at thousands of randomly selected L transformations arising in quartimax rotation. Not one of these was singular. The same was true for a smaller set of varimax rotations. There is presently, at least, no indication that our nonsingularity assumption will prove to be a practical difficulty. Indeed, in the cases looked at, the matrix (L_{ij}) was not only nonsingular but fairly well conditioned.

6. An Example and Discussion

Lawley and Maxwell [1971, p. 69] give an analysis of correlation coefficients between the scores of a sample of 292 children on a set of

10 cognitive tests. Their unrotated maximum likelihood estimates for loadings on three factors are given in Table 1. The standard errors of these loadings obtained by evaluating Lawley's formulas [Lawley and Maxwell, 1971, p. 62] for standardized loadings (i.e., loadings computed from correlations) are given in Table 2. These standard errors are pleasingly small, ranging in value from .022 to .094.

Insert Tables 1 and 2 about here

Turning to the rotated case, Table 3 contains a varimax rotation of the loadings in Table 1 together with the transformation matrix T which produced them [Lawley and Maxwell, p. 75]. As can be seen from the matrix T, a substantial rotation has been made. The resulting structure, however, is not particularly simple. The standard errors for the rotated loadings in Table 3 computed by using Lawley's formulas and ignoring the fact that T is computed from the data are given in Table 4. We will refer to these as uncorrected standard errors. Again their values are pleasingly small, ranging from .034 to .077. Using the results developed here, Table 5 contains the corresponding standard errors corrected for sample variation in T. The values cover roughly the same range, from .035 to .094. Before turning to a direct comparison of the uncorrected and corrected standard errors, we note that what has been presented thus far demonstrates the feasibility of computing standard errors for rotated loadings. To our knowledge this is the first time that standard errors for the rotated case have been presented in the literature.

Insert Tables 3, 4, and 5 about here

There are important and substantial differences between the uncorrected and corrected standard errors. Figure 1 is a plot of the cor-

----

Insert Figure 1 about here

----

rected standard errors of Table 5 against the uncorrected errors in Table 4. The uncorrected standard errors range from 41% below to 70% above the corrected standard errors. These differences support Wexler's [1968] simulation study which showed large discrepancies between uncorrected standard errors and standard errors obtained by simulation. The differences between uncorrected and corrected standard errors may be made arbitrarily large by choosing the data carefully. Using artificial data the authors have computed standard errors which differ by more than 50 fold. Because they originated from real data, however, the differences in Tables 4 and 5 are probably more relevant. It should be observed that the differences displayed in Figure 1 represent real theoretical discrepancies, not random fluctuations.

The standard errors presented give a simple indication of how stable factor loading estimates are. A quick significance test can be based on the rule which declares an observed difference significant if it exceeds twice the sum of the corresponding standard errors. Under this rule the

estimates in Table 3 of λ_{11} and λ_{12} differ significantly while those of λ_{12} and λ_{13} do not. A more sensitive test could be made by taking account of the covariances between the factor loading estimates.

Indeed, since the covariances are available, familiar procedures make it possible to test almost any hypothesis about rotated factor loadings. For example, is one variable related to the factors in the same way as another or, given a sample from a second population, does the factor pattern there differ from that of the present population? One may also produce simultaneous confidence intervals which allow him to scan across one or more tables of factor loadings in search of significant differences and account for the fact that he is scanning.

Before attacking such generalizations, however, it may be wise to ascertain how well the asymptotic results perform on finite samples. Preliminary work here is encouraging but only begun. Another natural next step is to derive similar results for the oblique case The results derived here could be used with other methods of extraction such as those of minres and alpha factor analysis except that the asymptotic covariances for unrotated loadings have not been derived for these methods. This provides still another area for investigation.

References

Girshick, M. A. On the sampling theory of roots of determinantal equations. Annals of Mathematical Statistics, 1939, 10, 203-224.

Harman, H. H. Modern factor analysis. Chicago: University of Chicago Press, 1967.

Lawley, D. N. A modified method of estimation in factor analysis and some large sample results. Uppsala Symposium on Psychological Factor Analysis. Nordisk Psykolige's Monograph Series No. 3. Stockholm: Almqvist and Wiksell, 1953. Pp. 35-42.

Lawley, D. N. Some new results in maximum likelihood factor analysis. Proceedings of the Royal Society of Edinburgh, 1967, 67A, 256-264.

Lawley, D. N., & Maxwell, A. E. Factor analysis as a statistical method. New York: American Elsevier, 1971.

Rao, C. R. Linear statistical inference. New York: Wiley, 1965.

Wexler, N. An evaluation of an asymptotic formula for factor loading variance by random methods. Unpublished doctoral dissertation. Rutgers University, 1968.

TABLE 1

Unrotated factor loadings for a set of cognitive tests

Test	Factor I	Factor II	Factor III	Communality
1. Comprehension	.788	-.152	-.352	.768
2. Arithmetic	.874	.381	.041	.911
3. Similarities	.814	-.043	-.213	.710
4. Vocabulary	.798	-.170	-.204	.707
5. Digit span	.641	.070	-.042	.418
6. Picture completion	.755	-.298	.067	.663
7. Picture arrangement	.782	-.221	.028	.661
8. Book design	.767	-.091	.358	.725
9. Object assembly	.733	-.384	.229	.737
10. Coding	.771	-.101	.071	.610

TABLE 2

Standard errors for the unrotated loadings in Table 1

	Factor		
Variate	I	II	III
1	.025	.094	.049
2	.040	.063	.046
3	.022	.083	.043
4	.024	.071	.048
5	.036	.076	.059
6	.028	.045	.059
7	.026	.049	.051
8	.027	.076	.051
9	.031	.059	.064
10	.026	.057	.048

TABLE 3

Varimax rotation of loadings in Table 1

	Factor			
Variate	I	II	III	Communality
1	.759	.329	.289	.768
2	.340	.849	.274	.911
3	.633	.450	.327	.710
4	.657	.345	.397	.707
5	.370	.453	.276	.418
6	.464	.263	.615	.663
7	.485	.332	.561	.660
8	.183	.472	.684	.725
9	.354	.209	.754	.737
10	.409	.423	.514	.610

Rotation Matrix T

.560	.633	.534
-.311	.758	-.573
-.768	.155	.622

TABLE 4

Uncorrected standard errors for the rotated loadings in Table 3

	Factor		
Variate	I	II	III
1	.038	.077	.067
2	.039	.064	.045
3	.038	.069	.055
4	.034	.060	.056
5	.057	.065	.055
6	.050	.046	.041
7	.042	.046	.042
8	.053	.065	.046
9	.063	.059	.034
10	.043	.051	.042

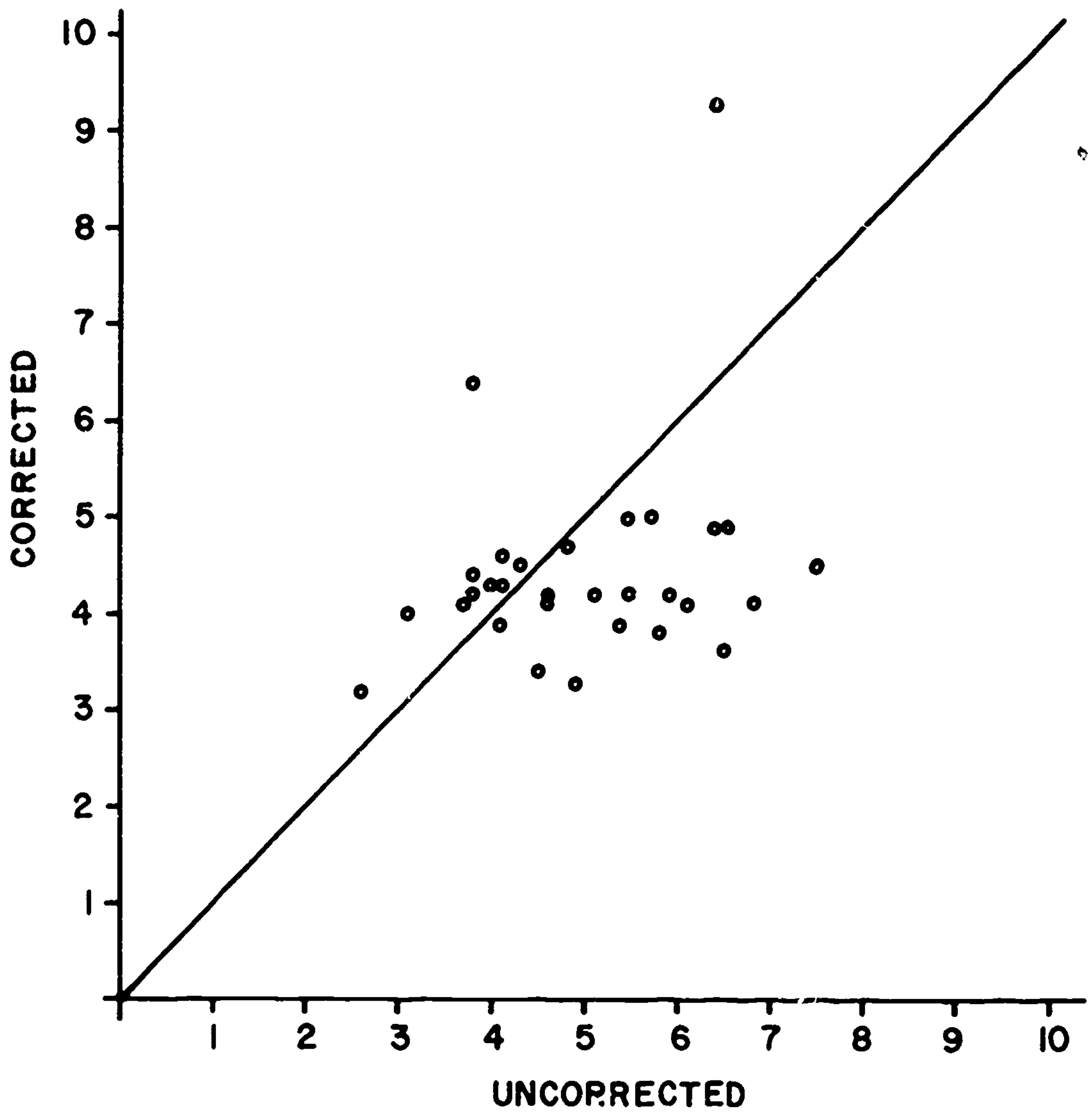
TABLE 5

Corrected standard errors for the rotated loadings in Table 3

Variate	Factor I	II	III
1	.041	.045	.036
2	.064	.094	.036
3	.044	.041	.039
4	.043	.043	.043
5	.051	.049	.050
6	.049	.044	.044
7	.048	.042	.043
8	.035	.050	.046
9	.047	.041	.042
10	.047	.043	.044

Figure Caption

Figure 1. Corrected verses uncorrected standard errors. Each corrected standard error in Table 5 is plotted against the corresponding uncorrected standard error from Table 4. A scale unit equals 0.01.


CORRECTED
UNCORRECTED
0
1
2
3
4
5
6
7
8
9
10